2018-10-04

Reaping the Benefits and Avoiding the Risks: Unrealistic Optimism in the Health Domain.

Hanoch, Y

http://hdl.handle.net/10026.1/12561

10.1111/risa.13204
Risk Anal

Reaping the Benefits and Avoiding the Risks: Unrealistic Optimism in the Health Domain

Yaniv Hanoch,[1,*] Jonathan Rolison,[2] and Alexandra M. Freund[3]

Abstract

People's perceptions of benefits and risks play a key role in their acceptance or rejection of medical interventions, yet these perceptions may be poorly calibrated. This online study with N = 373 adults aged 19 to 76 years focused on unrealistic optimism in the health domain. Participants indicated how likely they were to experience benefits and risks associated with medical conditions and completed objective and subjective numeracy scales. Participants exhibited optimistic views about the likelihood of experiencing the benefits and the side effects of treatment options described in the scenarios. Objective and subjective numeracy were not associated with more accurate ratings. Moreover, participants' underestimation of the risks was significantly greater than their overestimation of the benefits. From an applied perspective, these results suggest that clinicians may need to ensure that patients do not underestimate risks of medical interventions, and that they convey realistic expectations about the benefits that can be obtained with certain procedures.

Key Words: Risks, benefits, medical, numeracy, unrealistic optimism

1. Introduction

In his book, *Being Mortal: Medicine and what Matters in the End*, Atul Gawande (2014) discusses how terminally ill patients' and doctors' optimism about treatments and survival is often misguided, and how possible harms are often not mentioned or brushed aside. Supporting Gawande's insights, Jansen and colleagues (2011) found that participants in early-phase oncology trials—where typically *no* therapeutic benefits are expected—held unrealistic expectations about the trial outcomes, which was not due to misunderstanding the nature of the trial.

The Jansen et al study, like others, focused on patients' perception of the benefits, and failed to examine their beliefs about the risks. This is an important omission, as a systematic review (Hoffmann & Del Mar, 2015) reveals that patients not only overestimate the likelihood of experiencing benefits but that they also underestimate the possible risks associated with treatments, tests, and screenings. In fact, Hoffmann and Del Mar (2015) argue that even clinicians have relatively poor knowledge of risks, due in part, to the fact that risks are less often evaluated or even reported in primary research or reviews. Whether people overestimate the benefits and underestimate the potential harms has important clinical implications, e.g., for agreeing to take part in clinical trials, accepting vaccinations, or undergoing other medical procedures (see also Siegrist & Cvetkovich, 2001). Thus, it is important to understand how patients and health care providers view and define risks and benefits. The goal of the present study was to extend the scope of previous research to include both risks and benefits in the context of medical treatments as well as address a number of methodological shortcomings.

Inspired by work from decision theory, and endorsed by the European Medicine Agency, we employ the terms risks and benefits in this study to denote favorable effects for benefits and

(i) unfavorable effects or (ii) the likelihood of unfavorable effects for risks (Phillips, Fasolo, Zafiropoulos, & Beyer, 2011).

1.2 Unrealistic Optimism

In his seminal paper, Weinstein (1980) evaluated students' propensity to exhibit unrealistic optimism about a wide range of positive (e.g., traveling in Europe) and negative (e.g., developing cancer) life events. In this line of research (Weinstein, 1987; Weinstein & Nicolich, 1993), students are typically asked to rate how likely they believe they are to experience certain events compared to their fellow students. Results from these studies have repeatedly shown that students assume their chances of experiencing positive events to be higher and negative ones to be lower compared to their fellow students. Since Weinstein's (1980) original publication, investigators have extended this line of work to a wide range of domains. For example, following the 1989 California earthquake, Burger and Palmer (1989) showed that students were unrealistically optimistic about their likelihood of being hurt in a future earthquake. Other studies have demonstrated that motorcyclists (Rutter, Quine, & Albery, 1998) and car drivers (McKenna, 1983) exhibited unrealistic optimism regarding their chances of being involved in an accident. Likewise, smokers often underestimate their chances of developing serious health problems such as lung cancer (Weinstein, Marcus, & Moser, 2005) and heart disease (Ayanian & Cleary, 1999).

One area where unrealistic optimism has garnered particular attention is the health and medical domain. In fact, Weinstein and colleagues' earlier work (Weinstein, 1987; Weinstein & Nicolich, 1993) included many questions related to health conditions, such as the likelihood of developing cancer. A systematic review by Hoffmann and Del Mar (2015) identified thirty-five studies in which patients exhibited either unrealistically high expectations of a benefit or unrealistically low expectations of risk. For example, studies examining patients' perspectives on inflammatory bowel disease found that patients overestimated the benefits and underestimated the side effects of the associated medication (Baars, Markus, Kuipers, & Van Der Woude, 2010; Siegal, Levy, Mackenzie, & Sands, 2008). Other researchers have found that men and women overestimate the benefits of undergoing prostate and breast cancer screening and treatments, respectively (Domenighetti et al, 2003; Gigerenzer, Mata & Frank, 2009). Patients considering angioplasty are also overconfident that they will benefit from the treatment and not experience side effects (Habib, Sonoda, See & Groves, 2008).

There are potentially two issues with the studies reviewed by Hoffmann and Del Mar (2015). First, people might have difficulties conceptualizing and assessing risks for amorphous diseases such as cancer compared to precise risks (e.g., pain) or benefits (e.g., reduced fever). Second, many of the studies asked participants to provide an estimate of their chances without knowing the actual probabilities of benefits or risks. In a typical study, participants do not receive factual information about their actual chances of experiencing side effects or benefits; instead they are asked to base their assessment on their prior knowledge. In studies where participants receive relevant factual information, their responses tend to be much more in line with the factual information provided. In one illustrative study, Habib, Sonoda, See, Ell, and Groves (2008) furnished half of the patients with a risk assessment chart, detailing the risks and benefits. Those who consulted the chart made significantly more accurate judgments about the benefits and risks of the treatment compared to those who did not view the chart. These results nicely follow other studies showing that decision aids attenuate (but do not eliminate) patients' tendencies to exhibit an optimistic bias (Lewis, Pignone, Sheridan, Downs, & Kinsinger, 2003). Thus, results of the studies reviewed by Hoffmann and Del Mar might be partially due to participants' lack of relevant knowledge, which may have exaggerated an apparent unrealistic optimism.

Another line of criticism refers to concerns about the type of scales used to evaluate unrealistic optimism, which could distort the results (Harris & Hahn, 2011). An additional critique focuses on the sampling: due to the rare nature of events used in these studies, many participants will never experience them (for an extended response to these criticisms, see Shepperd, Klein, Waters, & Weinstein, 2013). In such cases, participants will not necessarily exhibit unrealistic optimism but an accurate reflection of their chances.

1.3 Numeracy and unrealistic optimism

Another factor that might mitigate participants' unrealistic optimism in the health domain is numeracy. A growing line of research has focused on the link between risk perception and numeracy, especially in the medical domain (Dieckmann, Slovic, & Peters, 2009; Gigerenzer, Gaissmaier, Kurz-Micke, Schwartz, & Woloshin, 2007; Kreuzmair, Siegrist, & Keller, 2017). Numeracy is generally understood as the capacity to comprehend and process numerical information (Reyna, Nelson, Han, & Dieckmann, 2009), as well as the ability to judge risk magnitude, compare risks, and estimate risk–benefit trade-offs. A large body of evidence indicates that persons with low numeracy skills have greater difficulties interpreting health risk information when presented in numerical format (Gigerenzer et al., 2007). Furthermore, low numerical ability is associated with misinterpretation of tests and screening results, such as mammography (Schwartz, Woloshin, Black, & Welch, 1997) and risk reductions associated with medical treatment options (Hanoch, Miron-Shatz, & Himmelstein, 2010). In a review of the literature, Reyna et al. (2009, p. 943) argued that "low numeracy distorts perceptions of the risks and benefits of screening, reduces medical compliance, impedes access to treatments, impairs risk communication (limiting prevention efforts among the most vulnerable), and …appears to adversely affect medical outcomes.” Surprisingly, no research thus far has examined the link between numeracy and unrealistic optimism.

In the present work, the use of both objective and subjective numeracy measures was motivated by an earlier critique of objective numeracy scales. Researchers have reported that objective numeracy scales often receive negative feedback from participants (Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, & Smith, 2007; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007). They also have poor completion rates and participants often object to assessments of their numerical ability. To circumvent these issues, we included the subjective numeracy scale (SNS) developed by Fagerlin and colleagues (2007). As a self-report measure of perceived numerical ability, the SNS tends to take less time to administer and is better suited to internet surveys due to its simplicity as participants are less likely to cheat by using a calculator or to drop out of the study if they feel that the test is too unpleasantly reminiscent of a math test in school. Studies have validated the SNS as a proxy for objective numeracy, demonstrating high correlations between SNS ratings and measures of objective numeracy using nationwide internet surveys (Zikmund-Fisher, et al., 2007), representative national samples (Galesic & Garcia-Retamero, 2010), and older adults (Rolison, Wood, Hanoch, & Liu, P-J, 2013).

1.4 Addressing earlier shortcomings

The present study tackles previous methodological shortcomings and investigates unrealistic optimism in the health domain. First, to address participants’ possible lack of knowledge, they were provided with precise risk information about each medical scenario. Second, participants estimated the chances that they would experience both positive (i.e., benefits) and negative (i.e., unwanted side effects) outcomes. To our knowledge, this is the first study that allows an assessment of which tendency—overestimating benefits or underestimating risks—is more pronounced. Third, participants were given exact information about specific benefits and risks (e.g., reduce fever, kidney failure), rather than a disease (e.g., cancer). To address previous concerns regarding scale attenuation, participants indicated their answers by moving a cursor on a scale ranging from 0%-100%. Finally, participants faced a wide range of realistic scenarios (including mental and dental conditions) and were asked to imagine that their doctor had recommended that they take a drug or undergo surgery. Taken together, the present study addresses the main methodological impediments identified in the literature.

The study tested the following hypotheses: (1) Participants overestimate the chances that they would experience the benefits and underestimate the chances that they would experience harm or side effects, indicating unrealistic optimism for benefits and risks; (2) Higher objective and subjective numeracy leads to more accurate judgments, thereby reducing unrealistic optimism; and (3) participants show greater optimism about not experiencing the harms than experiencing the benefits. Moreover, we included an age-diverse sample and included exploratory analyses of the association of age and unrealistic optimism.

2. Method

21. Procedure

Participants completed the study online. After providing informed consent, they were given a single practice trial to help familiarize them with the use of the scale. In the practice trial, all participants were asked to move a cursor on a scale ranging 0-100% and to place the cursor such that it indicated 40%. Only when they had accomplished this task could they move to the main study. Next, all health-related decision scenarios (both the ones that focused on benefits and the ones that focused on risk) were presented in random order to participants, such that the presentation of the scenarios was counterbalanced. Once participants completed the decision scenarios, they were asked to complete the objective and subjective numeracy scales. At the end of the survey, participants provided demographic information. The research protocol was in accordance with the ethics committee of the Department of Psychology at the University of Zurich.

2.2 Participants

We recruited MTurk participants with a HIT (human intelligence task) approval rate equal to or greater than 95%, and who were located in the United States. The results obtained through MTurk are reliable and comparable to those obtained by using hand-completed surveys (Casler, Bickel, & Hackett, 2013; Gibson, Piantadosi, & Fedorenko, 2011). Participants were reimbursed with U.S. $1.00. Participants identified their age in years and provided their year of birth at the beginning and at the end of the study. We removed 55 participants who either gave a year of birth that differed by more than one year from their reported age or provided two different years of birth. Some participants consistently provided either very low or very high ratings for side effects and benefits across the scenarios. We calculated the mean ratings across all side effect and benefit ratings and removed 43 outliers (i.e., who fell outside 95% of the data). Excluding these participants did not alter the pattern of results.

The final sample contained N = 373 participants, aged 19 to 76 years (M = 41.50, SD = 15.55; 48% female). Regarding education, one participant indicated that s/he had not completed high school, 11% indicated that they had completed high school, 28% had completed some college, 41% had completed college, and 16% indicated that they had completed a postgraduate (masters, doctoral) or professional degree. In the analyses that follow, we categorized education accordingly: 0 = high school or less; 1 = some college; 2 = vocational, technical, or college graduation; and 3 = completion of a postgraduate or professional degree. Regarding income, 10% indicated an annual household income under $15,000, 19% indicated an income between $15,000 and $30,000, 20% indicated an income between $30,001 and $45,000, 20%, indicated an income between $45,001 and $60,000, 11% indicated an income between $60,001 and $75,000, 9% indicated an income between $75,001 and $100,000, and 11% (*N* = 41), indicated an income that exceeded $100,000.

2.3 Materials

2.3.1 Medical scenarios: benefits. Participants reviewed five health scenarios that provided information about benefits. In these five scenarios, participants were asked to imagine that their doctor has recommended a treatment—a drug, dental surgery, ear surgery, kidney operation, or to take a newly developed medication—in order to treat an eye infection, a gum infection, a hole in their eardrum, a benign growth, and a life-threatening blood disorder, respectively. In each scenario, they were provided with precise information about the probability of success in both percentage and frequency formats (e.g., *"The probability of saving your tooth following the gum surgery is 25%-65%, that is, 25 to 65 out of 100 people who had the surgery retained their tooth*"). Participants were then asked to indicate how likely they believed that they were to experience one of the benefits (e.g., save their tooth) by moving a pointer on a scale from 0%-100%.

Participants were also asked to imagine that their doctor has recommended a treatment for a flu (a drug), a life threating illness (a drug), a fall (knee surgery), depression (anti-depression medication), and a heart problem (heart bypass surgery). They were then informed that not every person who accepts the treatment would experience the benefits. For example, in the heart bypass surgery scenario they were told that "*The probability that a person will* *experience one of the benefits is between 20% and 60%.*" Next, they were asked to indicate how likely they believed they personally were to experience five benefits—reduced risk of stroke, fewer problems with memory, reduced risk of death, reduced injury to the heart, and fewer heart rhythm problems—if they were to undergo the heart bypass surgery. Participants provided their responses by moving a pointer on a scale from 0% to 100%. In all scenarios, information about the benefits was taken from the medical literature, but the probability estimates were created for this study by the authors.

2.3.2 Medical scenarios: risks. Participants also answered questions about five health scenarios that were focused only on the risks. In these five scenarios, participants were asked to imagine that their doctor has recommended a treatment—a drug, dental surgery, skin surgery, back surgery, and ear drops. In each scenario they were provided with precise information about the probability of failure in both percentage and frequency formats (e.g., *The probability of being paralyzed following the back surgery is 30%-70%, that is, 30 to 70 people out of 100 who had this back surgery became paralyzed*). They were then asked to indicate how likely they believed they were to experience the associated harms (e.g., being paralyzed) by moving a pointer on a scale from 0% to 100%.

Moreover, participants were asked to imagine that their doctor has recommended a treatment for a life threating illness (a drug), an injury following a car accident (shoulder surgery), a general illness (vaccination), and a back problem (a spinal fusion). They were then informed that the treatment has several possible risks. For example, in the spinal fusion scenario they were told that *"The possibility that a person will experience one of the side effects is between 25%-65%*". Next, they were asked to indicate how likely they believed they were to experience the following five harmful outcomes or side effects (e.g., pain, failure of the fusion, blood clots, nerve injury, & infection) if they were to undergo the spinal fusion. Participants provided their responses by moving a pointer on a scale from 0% to 100%. In all scenarios, information about the risks was taken from the medical literature, but the probability estimates were created for this study by the authors.

2.3.3 Objective Numeracy was evaluated using an 11-item scale (Lipkus, Samsa & Rimer, 2001). The objective numeracy scale comprised 3 general questions of chance and probability (e.g. "*Out of 1,000 rolls, how many times do you think a fair, six-sided die would come up even*?") and 8 items related to understanding disease risk. Some of the disease risk items asked for an interpretation of risk (e.g. "*Which of the following represents the biggest risk of getting a disease? 1%, 5%, or 10%*"), others required that a percentage be converted into a frequency (e.g. "*If the chance of getting a disease is 10%, how many people would be expect to get the disease out of 100 people?"*), and some required that a frequency be converted into a percentage ("*If the chance of getting a disease is 20 out of 100, this would be the same as have a ___% chance of getting the disease*"). Responses were coded as either correct (1) or incorrect (0) and summed across the 11 items for overall numeracy scores.

2.3.4 Subjective numeracy was assessed using an 8-item subjective numeracy scale, in which participants are asked to report on their ability and preferences relating to statistical concepts on a 6-point scale (Fagerlin et al., 2007). Some items asked for ratings of ability in interpreting and calculating numerical values (e.g., "*How good are you at figuring out how much a shirt will cost if it is 25% off?*") ranging from "Not at all good" (with a numeric value of 1) to "Extremely good" (with a numeric value of 6). Other items asked for preferences for numeric over written information (e.g. "*When people tell you the chance of something happening, do you prefer that they use words [''it rarely happens''] or numbers [''there's a 1% chance'']?*"), ranging from “I always prefer words” (with a numeric value of 1) to “I always prefer numbers” (with a numeric value of 6). Responses were averaged across the 8 items for overall subjective numeracy.

2.3.5 Demographics Participants were asked to indicate their age, annual household income, education level, and gender.

3. Results

Participants had a mean objective numeracy score of 8.57 (*SD* = 1.88; out of 11) and a mean subjective numeracy score of 3.24 (*SD* = 0.80). Table 1 provides the inter-correlations and shows that higher objective numeracy was associated with higher subjective numeracy and higher educational attainment (Table 1).

Multiple conditions: Side effects. Provided in Figure 1 are participants’ mean ratings (Cronbach α = .76) for the side effect of each condition alongside the mid-point on the range provided in each scenario. In four of the five scenarios (except scenario 1; Figure 1), participants rated their personal risk as lower than the mid-point on the range provided in the scenario. On average, participants’ ratings were significantly below the mid-point of the range shown to them (M = -4.54, SD = 11.01, t[372] = 7.95, p < .001; Cohen’s d = 0.41), indicating optimistic views about the likelihood of experiencing side-effects. There were no significant correlations involving average ratings across scenarios (see Table 1).

Multiple side effects. Figure 2 shows participants’ mean ratings (Cronbach α = .76) for the multiple side effects of one of the scenarios. As in all scenarios, participants underestimated their personal risk with respect to the mid-point of the percentage range they were shown. We conducted our statistical analysis on average ratings for the five side effects of each of the five scenarios that presented multiple side effects. A one-way repeated measures analysis of variance

(ANOVA) revealed a significant effect of scenario on participants' mean ratings across side-effects ($F(4,1488) = 198.82$, $p < .001$, $\eta^2 = .35$), such that participants' underestimation of their personal risk was greatest for side effects of drug prescription for a headache ($M = -22.04$; $SD = 17.12$), followed by a drug prescription for a life-threatening illness ($M = -11.98$; $SD = 13.79$), a recommendation to undergo spinal fusion ($M = -11.36$; $SD = 12.63$), a vaccination ($M = -10.71$; $SD = 14.40$), and surgery following a car accident ($M = -5.69$; $SD = 12.19$). Paired-samples *t*-tests confirmed significant differences between side effects of drug prescription for a headache and a drug prescription for a life-threatening illness ($t[372] = 17.58$, $p < .001$) and between a vaccination and surgery following a car accident ($t[372] = 9.23$, $p < .001$). These differences may partly reflect the position of the mid-point on the percentage range as this was highest for the scenario describing a drug prescription for a headache (mid-point = 60) and was lowest for the scenario describing surgery following a car accident (mid-point = 45). After adjusting for the position of the scale mid-point by dividing the difference scores by the respective mid-point value, the effect of scenario reduced in size ($F[4,1488] = 45.36$, $p < .001$, $\eta^2 = .11$). There were no significant correlations involving participants' mean ratings across the five scenarios with other variables other than with their mean ratings for the side effects for the multiple conditions scenario (see Table 1).

Multiple conditions: Benefits. Figure 3 shows participants' mean ratings (Cronbach $\alpha = .86$) for the benefit associated with each scenario alongside the mid-point of each percentage range provided to participants. The personal benefits reported by participants for each medical scenario exceeded the mid-point on the percentage range they were shown. On average, participants' ratings were significantly above the mid-point of the range shown to them ($M = 4.17$, $SD = 12.45$, $t(372) = 6.47$, $p < .001$), indicating optimistic views about the likelihood of experiencing benefits associated with the scenarios. Ratings were not significantly associated with age or objective numeracy, but were positively associated with subjective numeracy, educational attainment, and income (Table 1). Ratings for the likelihood of benefits were also negatively associated with ratings for the risk of side effects (Table 1), indicating that greater optimism about the potential benefits was accompanied by greater optimism about the potential risks.

Multiple benefits. Figure 4 provides participants' mean ratings (Cronbach α = .86) for the multiple benefits of one of the scenarios. As in all scenarios, participants overestimated the likelihood of experiencing the benefits with respect to the mid-point of the percentage range they were shown. We conducted our statistical analysis on average ratings for the five benefits of each of the five scenarios that presented multiple side effects. A one-way repeated measures analysis of variance revealed a significant effect of scenario on participants' mean ratings across benefits ($F[4,1488] = 117.73$, $p < .001$, $\eta^2 = .24$), whereby participants overestimated their personal likelihood of benefit by the greatest amount for a recommendation to undergo a heart bypass surgery ($M = 8.68$; $SD = 14.77$), followed a recommendation to undergo a knee surgery following a fall ($M = 7.72$; $SD = 14.84$), a drug for the flu ($M = 3.14$; $SD = 13.61$), and a drug for a life-threatening illness ($M = 0.99$; $SD = 12.99$). For a scenario describing a recommendation for anti-depression medication participants instead underestimated their likelihood of benefit ($M = -2.84$; $SD = 14.35$). Paired-samples *t*-tests confirmed significant differences between benefits to undergo a heart bypass surgery and to undergo a knee surgery following a fall ($t[372] = 1.88$, $p = .048$), between benefits to undergo a knee surgery following a fall and a drug for the flu ($t[372] = 7.20$, $p < .001$), between benefits of a drug for the flu and a drug for a life-threatening illness ($t[372] = 3.67$, $p < .001$), and between a drug for a life-threatening illness and an anti-depression medication ($t[372] = 6.23$, $p < .001$). As with side-effects, these differences may partly reflect the mid-point on the percentage range, which was lowest in the scenario describing a recommendation to undergo a heart bypass surgery (mid-point = 40) and was highest in the scenario that described a drug for a life-threatening illness (mid-point = 60). After adjusting for the position of the scale mid-point by dividing the absolute difference scores by the respective mid-point value, the effect of scenario reduced in size ($F[4,1488] = 54.08$, $p < .001$, $\eta^2 = .13$). Mean self-ratings across the five scenarios were not significantly associated with age or objective numeracy, but were positively associated with subjective numeracy, education, and income. Moreover, higher mean ratings for the likelihood of benefits were associated with higher likelihood ratings of benefits for the multiple conditions scenario and with lower ratings for the risk of side effects (see Table 1).

Calibration. We assessed whether demographic variables and objective and subjective numeracy were associated with the degree to which participants' ratings were calibrated with the mid-point of the percentage range. This was done by calculating the mean absolute difference between participants' ratings and the mid-point of the scale across side effects and benefits scenarios. Lower scores indicate a lower mean absolute difference between a participant's rating and the scale mid-point across side effects and benefits scenarios, which indicates better calibration. Calibration scores were calculated separately for the multiple conditions and multiple side effects/benefits scenarios for each participant. This analysis showed that calibration scores did not correlate significantly with age, education, income, objective numeracy, or subjective numeracy (see Table 1).

Underestimation of side-effects versus overestimation of benefits. To assess whether participants underestimated the side-effects to a greater degree than they overestimated the benefits, we reverse scored participants' ratings for the side-effects scenarios to enable comparison in ratings for side-effects and benefits. As our previous analyses suggested that participants were sensitive to the location of the mid-point on the percentage range, we compared scenarios according to their mid-points. When the mid-point of the percentage range was 50%, participants underestimated the risks (M = 10.71) to a greater degree than they overestimated the benefits (M = 3.14; t[372] = 8.25, $p < .001$). When the mid-point for the percentage risk of side-effects was 60% and the mid-point for the percentage likelihood of benefit was 40%, allowing the same range of under- and over-estimation, respectively, underestimation of the risks (M = 22.04) was greater than overestimation of the benefits (M = 8.68; t[372] = 14.33, $p < .001$). Underestimation of the risk was also significantly greater than overestimation of the benefits when the mid-point for side-effects was 45% and the mid-point for benefits was 55% (M = 11.98 vs. M = -2.84; t[372] = 15.83, $p < .001$; M = 11.36 vs. M = -2.84; t[372] = 16.12, $p < .001$) and when the mid-point for side-effects was 40% and the mid-point for benefits was 60% (M = 5.69 vs. M = 0.99; t[372] = 5.84, $p < .001$).

4. Discussion

Two of the key factors involved in accepting or rejecting medical interventions are the benefits and the risks associated with them. Furthermore, understanding the risks and benefits of any intervention is crucial for providing informed consent (Ross, 2015). Yet, as the systemic review by Hoffmann and Del Mar (2015) has shown, patients seldom have an accurate representation of the benefits and risks of medical options. Patients often overestimate their likelihood of experiencing the benefits and underestimate their likelihood of experiencing the risks. By presenting participants with a wide range of medical scenarios—including minor and serious ones, as well as physical, psychological and dental—our findings lend support to a growing corpus of evidence regarding unrealistic optimism. Moreover, this study addressed many of the methodological concerns raised in earlier work. Crucially, as one of the few studies to include questions about both benefits and risks, our data allowed us to assess whether people are more optimistic about risks or benefits. Results demonstrate that participants were more optimistic about their chances of avoiding the risks than reaping the benefits. Aside from its novelty, this finding is also especially notable considering earlier work (Hoffmann & Del Mar, 2015), indicating that health care professionals are less likely to discuss risks than benefits with their patients.

One criticism of earlier studies is the lack of precise information provided to participants and patients in making their estimates. This line of reasoning seems intuitively reasonable, and studies that provided patients with additional information indeed saw a reduction in unrealistic optimism. Participants in our study received precise risk and benefit information as a percentage range and in frequency format to facilitate their comprehension. However, in contrast to earlier findings, this did not impact participants' judgments (Gigerenzer et al, 2009). One possible reason for the divergent results could rest with the fact that earlier studies (Habib et al., 2008) provided participants with decision aids rather than with additional information. It is also possible that our use of a range (e.g., 20%-60%) rather than a single number (e.g., 40%) affected the results, as lay people (and professionals) often encounter difficulties understanding numerical information presented as a range (Dieckmann, et al., 2009; Dieckmann, Peters, & Gregory, 2015). Hanoch, Miron-Shatz, and Himmelstein (2010), for example, have reported that women as well as health care providers often misunderstand lifetime risk of developing breast cancer when present as a range between 35-85% (for similar results with men see Rolison, Hanoch & Miron-Shatz, 2012). In cases where a single risk number is provided (e.g., 40%), it might be harder for individuals to rate their personal risk differently than the communicated number. However, when a range is provided, persons have more room to exhibit unrealistic optimism. Our usage of a range (rather than single number) was motivated by the fact that many medical risks and benefits of procedures such as genetic testing and lifetime risk are represented along a continuum rather than as a single probability, thereby increasing the external validity of our method. In our study, participants imagined medical scenarios (e.g., "*The probability of saving your tooth following the gum surgery is 25%-65%, that is, 25 to 65 out of 100 people who had the surgery retained their tooth*"). In real clinical practice, however, patients are likely also to possess additional medical information (e.g., clinical test results) that better inform them about their actual risk.

Our study was also designed to address concerns regarding peoples' ability to think about a general condition (such as cancer) rather than specific symptoms (e.g., pain). Weinstein's studies typically asked students to compare their probabilities of developing cancer to their fellow students. In contrast, we asked participants about specific and realistic outcomes (e.g., fever following a surgery). Indeed, we developed our materials based on actual benefits and risks that one could expect from the actual interventions. Furthermore, we also included very common conditions (such as flu) to circumvent concern regarding the rarity and prevalence of the condition. However, providing participants with specific rather than general information, and with common rather than rare conditions, seems to do little to reduce unrealistic optimism.

Harris and Hahn (2011) have argued that the type of scale used to measure unrealistic optimism distorts study results. We addressed this criticism by requiring participants to respond by moving a cursor on a continuous ruler from 0%-100% rather than asking them to indicate on a Likert scale (e.g., from -3 to +3). Using rulers, such as the one utilized in this study, improves risk comprehension and communication (Van Belle & Van Calster, 2015). Thus, if Harris and Hahn's intuition were correct, one would expect the overestimation of benefits and the underestimation of risks to be reduced or even to disappear with the ruler response format. Our data speak against this expectation: Using a ruler, where participants had to move the cursor to indicate their response rather than a numerical response scale, did little to change earlier findings.

We expected higher numeracy to be related to lower unrealistic optimism. There is now a large body of evidence showing the importance of numeracy in a wide spectrum of medical decisions (Gigerenzer et al., 2007; Reyna, et al., 2009), showing also that higher numeracy is related to improved risk comprehension (Hanoch, et al., 2010). Our data did not follow this pattern. Rather, numeracy levels (whether objective or subjective) were largely unrelated to participants' responses. This finding is in line with an earlier study (Hanoch, Rolison, & Freund, 2018) that also found no link between numeracy and medical risk taking. Thus, our finding suggests that improving numeracy skills might not help solve the problem of misjudging risks and benefits of medical procedures. To address this problem, other means need to be developed, such as ensuring the health care providers discuss risk information with their patients and developing decisions aids (se.g., Habib et al., 2008; Lewis, et al., 2003) that are specifically designed to negate unrealistic optimism.

There are several possible explanations for the lack of association between numeracy and risk/benefit estimates. First, numeracy might offer little protection against the psychological processes related to optimism, as people may be motivated to believe that they have a low probability of experiencing the negative outcome irrespective of their numeracy skills. Second, numeracy might be more closely associated with comprehension than with the perception of benefits and risks, which was the aim of the current study. Finally, our results could also be driven by the fact that our sample was highly numerate, both objectively and subjectively, reducing the variance within the numeracy variables. A more diverse sample regarding numeracy might provide a different picture.

Our finding that subjective but not objective numeracy was related to perceptions of benefits resonates with research suggesting that risk assessment is not performed solely through cognitive paths but is evaluated via what has been termed “risk as feeling” (Keller, Siegrist, & Gutscher, 2006; Lowenstein, Weber, Hsee, & Welch, 2001). Furthermore, Peters and colleagues (2006) have argued that people high in numeracy might even be more biased than those lower in numeracy when assessing risk information, although the authors found this in one type of risk judgment and it may not generalize to other risk domains. This could be the case, as more numerate people tend to rely more heavily on the affective meaning of numbers, which leads them to worse decisions. A study by Miron-Shatz, Hanoch, Doniger, Omer, and Ozanne (2014) with high risk women provides further support to this interpretation, showing that subjective numeracy, but not objective numeracy, is related to women’s willingness to pay for genetic testing. In addition, Liberali and colleagues (2012) have suggested that, although correlated, subjective and objective numeracy measure different constructs.

Despite our promising results, several limitations need to be discussed. First, our study was conducted with a sample drawn from the general population and not with patients. It is possible that similar studies with patient populations, who could actually benefit from consultation with their health care provider, would provide different results. However, previous studies (Hoffmann & Del Mar, 2015) cast doubt on this possibility. Furthermore, our study is based on self-reports, and it is unclear whether they can be generalized to actual medical consultations. Using a probability scale (one that ranges from 0-100%) could have influenced our results, as researchers have argued that using such scales can lead to inflated use of the 50% response (Bruine de Bruin, Fischbeck, Stiber, & Fischoff, 2002). In future investigations researchers should be cognizant of this issue and follow recommendations of Bruine de Bruin et al. (2002). Finally, we used Mturk to recruit our sample. While this platform is an excellent recruitment resource, the demographics of this sample profile is not representative for the general population.

What are the implications of our results for practitioners? As noted above, an accurate representation of risks and benefits is one of the hallmarks of informed consent. Thus, clinicians need to ensure that patients do not under- or over-estimate the success or failure of any given intervention. Likewise, patients' willingness to accept or request interventions is driven by their assumptions regarding the therapeutic benefits and risks of these interventions. Furthermore, as Atul Gawande's book powerfully demonstrates, many patients' unrealistic optimism may lead to unnecessary, unneeded, and often counterproductive or even harmful interventions (Moynihan, Bero, Ross-Degnan, et al., 2000). In addition, unrealistic expectations on the patients' side have been identified as one possible culprit in the rise of medical costs (Lipitz-Snyderman & Back, 2013). Taken together, ensuring that people's perception of medical interventions (e.g., tests, procedures and treatments) is accurate could potentially improve clinical outcomes as well as reduce economic burden on the health-care system.

References

Ayanian, J. Z., & Cleary, P. D. (1999). Perceived risks of heart disease and cancer among cigarette smokers. *JAMA,* 81, 1019–1021.

Baars, J. E., Markus, T., Kuipers, E. J., *&* Van Der Woude, C. J. (2010). Patients' preferences regarding shared decision-making in the treatment of inflammatory bowel disease: results from a patient-empowerment study. *Digestion*, 81, 113–9

Bruine de Bruin, W., Fischbeck, P. S., Stiber, N. A., & Fischhoff, B. (2002). What number is 'fifty-fifty'? redistributing excessive 50% responses in elicited probabilities. *Risk Analysis, 22*(4), 713–723.

Burger, J. M, & Palmer, M. L. (1992). Changes in and generalization of unrealistic optimism following experiences with stressful events: Reactions to the 1989 California earthquake. *Personality and Social Psychology Bulletin*, 18, 39–43. doi:10.1177/0146167292181006

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computer and Human Behavior, 29*(6), 2156–2160.

Dieckmann, N. F., Slovic, P., & Peters, E. M. (2009). The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. Risk Analysis, 29, 1473–1487. DOI: 10.1111/j.1539-6924.2009.01279

Dieckmann, N. F., Peters, E., and Gregory, R. (2015). At home on the range? Lay interpretations of numerical uncertainty ranges. *Risk Analysis,* 35, 1281–1295. doi: 10.1111/risa.12358Domenighetti

G, D'Avanzo B, Egger M, Berrino F, Perneger T, Mosconi P, Zwahlen M. Women's perception of the benefits of mammography screening: population-based survey in four countries. *International Journal of Epidemiology.* 2003; 32(5): 816-821.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, 27(5), 672–680.

Galesic, M, & Garcia-Retamero, R. (2010). Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Archive of Internal Medicine*, 170, 462-468.

Gawande A. *Being Mortal: Medicine and what Matters in the End*. 2014. Profile Books.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using mechanical turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, *5*(8), 509–524. doi:10.1111/j.1749-818X.2011.00295.x

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96. doi:10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of National Cancer Institute*, 101(17), 1216-1220.

Habib, S. B., Sonoda, L., See, T. C., Ell, P. J., & Groves, A. M. (2008). How do patients perceive the benefits and risks of peripheral angioplasty? Implications for informed consent. *Journal of Vascular Interventional Radiography*, 19, 177-81.

Hanoch, Y., Miron-Shatz, T., & Himmelstein M. (2010). Genetic testing and risk interpretation: How do women understand lifetime risk results? *Judgment and Decision Making*, 5, 116-123.

Hanoch, Y., Rolison, J., & Freund A. M. (2018). Does Medical Risk Perception and Risk Taking Change With Age? Risk Analysis, 38, 917-928.

Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. *Psychological Review*, 18, 135–154. doi:10.1037/a0020997

Hoffmann, T., & Del Mar, C. (2015) Patient expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Internal Medicine*, 175, 274–86.

Jansen, L. A., Appelbaum, P. S., Klein, W. M. P., Weinstein, N. D., Cook, W., Fogel, J. S., & Sulmasy, D. P. (2011). Unrealistic optimism in early-phase oncology trials. *IRB: Ethics & Human Research*, 33(1), 1–8.

Keller, C., Siegrist, M., & Gutscher, H. (2006). The role of the affect and availability heuristic in risk communication. Risk Analysis, 26, 631–639. doi:10.1111/j.1539-6924.2006.00773.x

Kreuzmair, C., Siegrist, M. & Keller, C. (2017). Does iconicity in pictographs matter? The influence of iconicity and numeracy on information processing, decision making, and liking in an Eye-Tracking study. Risk Analysis, 37, 546-556.

Lewis, C. L., Pignone, M. P., Sheridan, S. L., Downs, S. M., & Kinsinger, L. S. (2003). A randomized trial of three videos that differ in the framing of information about mammography in women 40 to 49 years old. *Journal of General Internal Medicine*, 18, 875–883

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25*, 361–381.

Lipitz-Snyderman, A., & Bach P. B. (2013). Overuse of health care services: when less is more … more or less. *JAMA Internal Medicine*, 173(14), 1277-1278.

Lipkus, M, Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37–44.

Lowenstein, G. E., Weber, E.U., Hsee, C. K., & Welch, N. (2001). Risk as feeling. *Psychological Bulletin, 127,* 267–286.

McKenna, F. M. (1993). It won't happen to me: Unrealistic optimism or illusion of control? *British Journal of Psychology,* 84, 39–50.

Miron-Shatz, T., Hanoch, Y., Doniger, G. M., Omer, Z. B. & Ozanne, E. M. (2014). Subjective but not objective numeracy influences willingness to pay for BRCA1/2 genetic testing. Judgment and Decision Making, 9, (2) 152-158

Moynihan, R, Bero, L, Ross-Degnan, D, et al. (2000). Coverage by the news media of the benefits and risks of medications. *New England journal of Medicine,* 342(22), 645-1650.

Peters E, Västfjäll D, Slovic P, Mertz C, Mazzocco K, Dickert S. (2006). Numeracy and decision making. *Psychological Science*, 7(5), 407–413.

Phillips, L. D., Fasolo, B., Zafiropoulos, N., & Beyer, A. (2011). Is quantitative benefit–risk modelling of drugs desirable or possible? *Drug Discovery Today: Technologies,* 8 (1). e3-e10. ISSN 1740-6749

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135, 943–973. doi:10.1037/a0017327

Rolison, J. J., Hanoch, Y., & Miron-Shatz, T. (2012). What do men understand about lifetime risk following genetic testing? The effect of context and numeracy. *Health Psychology*, 31, (4) 530-533

Rolison, J. J., Wood, S., Hanoch, Y., & Liu, P-J. (2013). Subjective numeracy scale as a tool for assessing statistical numeracy in older adult populations. *Gerontology*, 59, (3) 283-288.

Rolison, J. J., Morsanyi, K., & O'Connor, P. A. (2016). Can I count on getting better? Association between math anxiety and poorer understanding of medical risk reductions. *Medical Decision Making,* 36, 876–886. doi.org/10.1177/0272989X15602000

Ross, J. S. (2015). Informed consent and communicating risk and benefits of research on higher-risk medications. *JAMA Internal Medicine,* 175, 1568–1569. doi:10.1001/jamainternmed.2015.3555

Rutter, D. R., Quine, L., & Albery, I. P. (1998). Perceptions of risk in motorcyclists: Unrealistic optimism, relative realism and predictions of behaviour. *British Journal of Psychology,* 89, 681–696.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal Medicine,* 127, 966–972.

Siegel, C. A., Levy, L, C., Mackenzie, T. A. (2018). Sands BE. Patient perceptions of the risks and benefits of infliximab for the treatment of inflammatory bowel disease. *Inflammatory Bowel Disease*, 14(1),= 1-6.

Siegrist, M., & Cvetkovich, G. (2001). Better negative than positive? Evidence of a bias for negative information about possible health dangers. Risk Analysis, 21, 199 – 206.

Shepperd, J. A., Klein, W. M. P., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspective on Psychological Science*, 8, 395–411. doi:10.1177/1745691613485247

Van Belle, V., Van Calster, B. (2015). Visualizing Risk Prediction Models. *PLoS ONE,* 5: 10(7):e0132614. pmid:26176945

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology,* 39, 806–820.

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine,* 10, 481–500. doi:10.1007/bf0084614.

Weinstein, N. D., Marcus, S. E., Moser, R. P. (2005). Smokers' unrealistic optimism about their risk. *Tobacco Control,* 14, 55–59.

Weinstein, N. D., & Nicolich, M. (1993). Correct and incorrect interpretations of correlations between risk perceptions and risk behaviors. *Health Psychology*, 12, 235–245. doi:10.1037/0278-6133.12.3.235

Zikmund-Fisher, B. J., Smith, D. M, Ubel, P. A., & Fagerlin A. (2007). Validation of the Subjective Numeracy Scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making*, 27, 663-671.

Table 1. *Inter-correlations between demographic variables, objective and subjective numeracy, and responses to medical scenarios.*

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Age (1)	–	.04	.08	.10	.03	.06	.08	.05	-.01	-.04	-.03	.06	-.04
Education (2)		–	.31**	.20**	.23**	.05	.00	.11*	.12*	.09	.10	.09	.05
Income (3)			–	.10	.15**	.03	-.07	.10*	.11*	.07	.02	.08	.05
Objective numeracy (4)				–	.45**	.01	.02	.09	.06	-.04	-.11*	-.07	-.05
Subjective numeracy (5)					–	-.06	-.01	.12*	.13*	-.05	-.11*	-.03	-.06
Multiple conditions: Side effects (6)						–	.52**	-.15**	-.36**	.12*	-.04	-.35**	-.33**
Multiple side effects (7)							–	-.29**	-.37**	-.35**	-.40**	-.50**	-.75**
Multiple conditions: Benefits (8)								–	.62**	.28**	.24**	.56**	.45**
Multiple benefits (9)									–	.12*	.14**	.52**	.44**
Response consistency: Side effects (10)										–	.63**	.47**	.64**
Response consistency: Benefits (11)											–	.54**	.71**
Calibration: Multiple conditions (12)												–	.77*
Calibration: side effects / benefits (13)													–

* $p \leq .05$; ** $p \leq .01$;

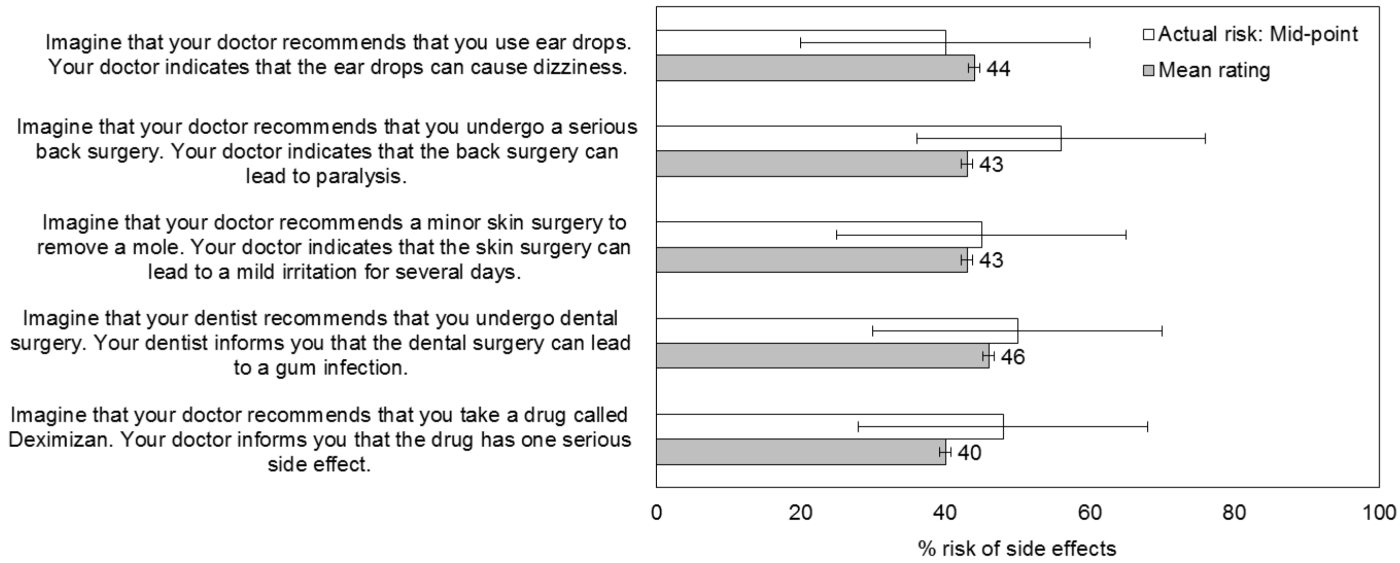

Figure 1. Actual risk of side effects as the mid-point on the range provided in the scenario and the mean ratings provided by participants for each of the side effects scenarios. The horizontal bars represent 1 standard error either side of the mean rating. The horizontal bars around the mid-point indicate the upper and lower bounds provided in scenario.

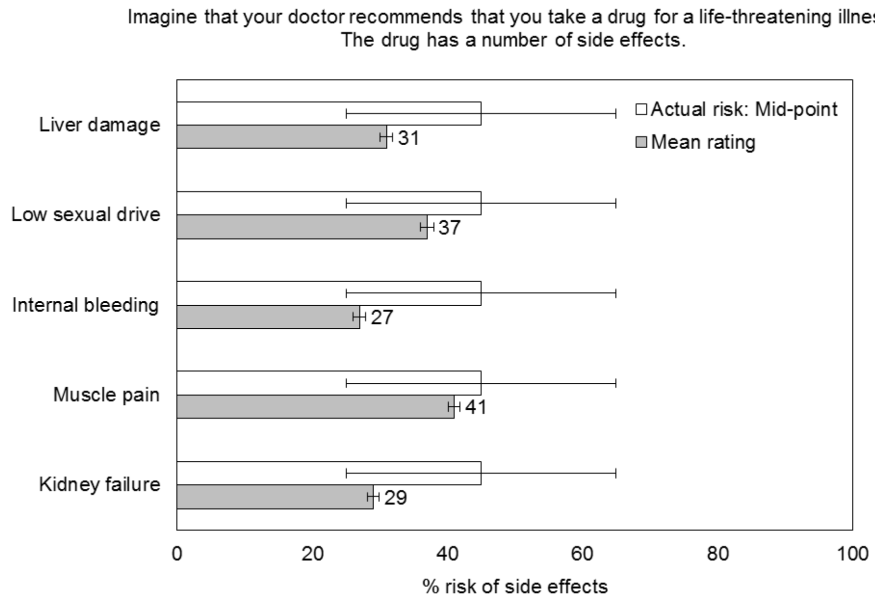

Figure 2. Actual risk of side effects as the mid-point on the range provided in the first multiple side effects scenario and the mean ratings provided by participants for each of the multiple side effects. The horizontal bars represent 1 standard error either side of the mean ratings. The horizontal bars around the mid-point indicate the upper and lower bounds provided in scenario.

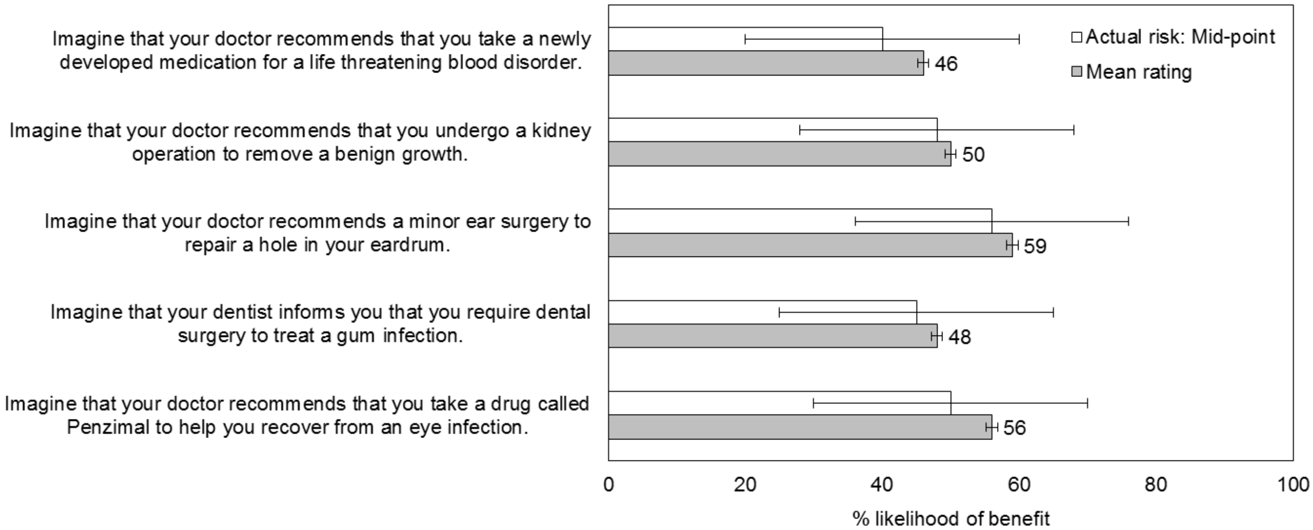

Figure 3. Actual chances of benefit as the mid-point on the range provided in the scenario and the mean ratings provided by participants for each of the benefits scenarios. The horizontal bars around the mean ratings represent 1 standard error either side of the mean rating ratings. The horizontal bars around the mid-point indicate the upper and lower bounds provided in scenario.

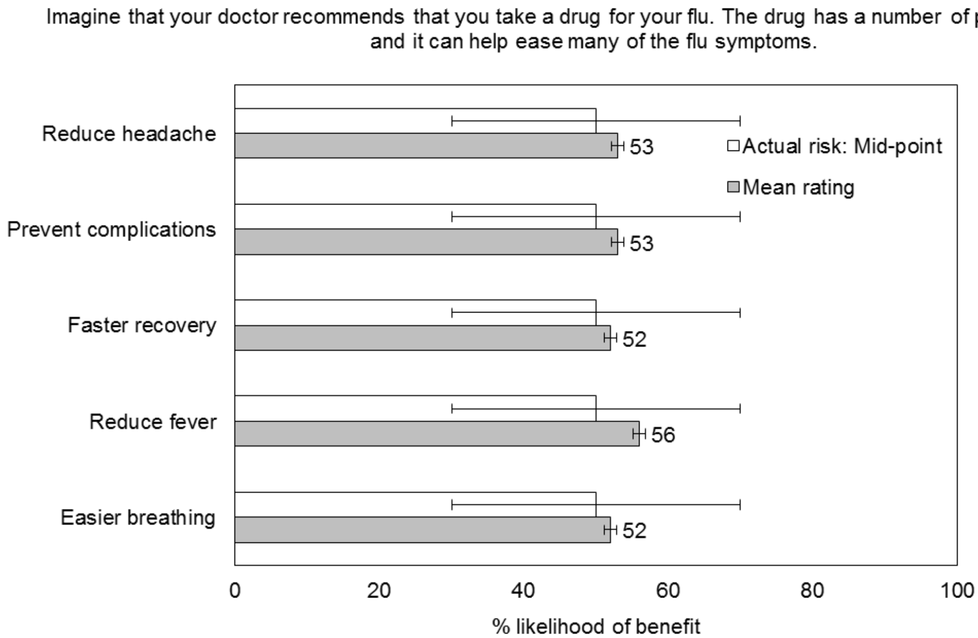

Figure 4. Actual chances of benefit as the mid-point on the range provided in the scenario and the mean ratings provided by participants for each of the benefits. The horizontal bars around the mean ratings represent 1 standard error either side of the mean rating ratings. The horizontal bars around the mid-point indicate the upper and lower bounds provided in scenario.